How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences

Ralph Adolphs

Division of Humanities and Social Sciences
California Institute of Technology

radolphs@caltech.edu
HSS 228-77, Caltech
Pasadena, CA 91125

Running title: Affective neuroscience needs distinctions

Abstract.

In this debate with Lisa Feldman Barrett, I defend a view of emotions as biological functional states. Affective neuroscience studies emotions in this sense, but it also studies the conscious experience of emotion ("feelings"), our ability to attribute emotions to others and to animals ("attribution", "anthropomorphizing"), our ability to think and talk about emotion ("concepts of emotion", "semantic knowledge of emotion"), and the behaviors caused by an emotion ("expression of emotions", "emotional reactions"). I think that the most pressing challenge facing affective neuroscience is the need to carefully distinguish between these distinct aspects of "emotion". I view emotion states as evolved functional states that regulate complex behavior, in both people and animals, in response to challenges that instantiate recurrent environmental themes. These functional states, in turn, can also cause conscious experiences (feelings), and their effects and our memories for those effects also contribute to our semantic knowledge of emotions (concepts). Cross-species studies, dissociations in neurological and psychiatric patients, and more ecologically valid neuroimaging designs should be used to partly separate these different phenomena.

Keywords: emotion, concepts, affective neuroscience, amygdala, feelings

Words in main text: 4808

The neuroscientific investigation of emotion, affective neuroscience, is one of the most interesting and vibrant new disciplines, and of paramount importance for understanding individual differences and many psychiatric disorders. It is also one of the most confusing disciplines, in large part because the word "emotion" is used in multiple ways. In this debate with Lisa Feldman Barrett, we will try to make it less confusing. I begin with a bare-bones decription of my view (cf. **Figure 1**), then discuss how confusions arise, and conclude with some thoughts on how the neuroscientific investigation of emotions could move forward. Since my focus is on how even to think about "emotion" as scientists, this article is light on reviews of empirical studies and heavier on conceptual material. I aim here to present my own view as clearly as possible, and will reserve most of my comments on Lisa's view for the rebuttal.

Affective neuroscience, and cognitive neuroscience more generally, requires a close interplay between the vocabularies and frameworks of two different scientific disciplines: psychology and neuroscience. When they study emotions, these two disciplines are trying to describe the same states or processes (I use these terms interchangeably here), but often on the basis of different kinds of data, methods, and theories. Some views suppose one discipline has priority over the other, often that neuroscience trumps psychology: if we can't find evidence for a psychological construct from brain data, then the psychologists were wrong about that construct. I want to resist such a view, in large part because we really have very little idea about how to interpret neuroscience data, so whatever evidence it does or does not provide for a psychological theory should be considered extremely preliminary. So while I believe that emotions are brain states, I also believe that we need to begin by understanding them as psychological states.

1. What are emotions? Emotions are functional states, implemented in the activity of neural systems, that regulate complex behaviors.

My view of emotions begins with everyday observations and concepts (but it doesn't end there). Clear instances of an emotion state are those that Charles Darwin already noted in his book, *The Expression of the Emotions in Man and Animals* [1]. That is, at least in the first instance, we attribute emotion states to people and to many other animals on the basis of their context-dependent observed behavior. Emotion states, together with many other mental state attributions, provide causal explanations of behavior. I take it that there is little disagreement that a weeping adult, a screaming child, and a hissing cat all are in various emotion states. What exactly we want to call those states, and whether they should be thought of categorically or dimensionally, are further and more complicated questions. For simplicity of exposition I will use words like "fear" to refer to the hissing cat in this paper, for example-- but the meaning of that term will likely need to be revised to be useful in a mature affective neuroscience. It is also the case that the sets of stimuli and behaviors that specify an emotion's functional role are highly context-dependent. For instance, the cat's hissing behavior by itself is also quite compatible with an emotion state of anger, rather than fear (or both) -- one would need to do experiments (challenging the animal with specific stimuli, observing its behavior) to disambiguate this, since the hissing is not unique to fear. Likewise, the

weeping adult may be weeping from sadness, or anger, or some combination of these emotion states as we commonly conceive them-- we would need more information about their facial expression, their other behavior, and the circumstances in which the behavior is observed. This applies to most emotional behaviors: they help narrow down the set of possible emotion states that explain the behavior, but we typically require considerable additional information about the history and circumstances under which the behavior is exhibited to disambiguate between multiple possible emotion states that could account for the behavior. We do this all the time: if somebody is behaving in a way indicative of an emotion state, we usually probe them further to get additional evidence (for instance, by asking them how they feel, or what happened, or what they intend to do). The context-dependency of emotion states is also critical to consider for affective neuroscience studies in which we want to experimentally manipulate emotion states.

If the above simple starting point is a reasonable place to begin to develop a scientific concept of emotion states, the next question is what it is about these examples -- the weeping person, screaming child, or hissing cat-- that distinguish them as evidence for emotion states. All behaviors are caused by central states of various kinds: so what distinguishes emotion states? A useful comparison is with behaviors that are either simpler or more complex. Reflexes are simpler than emotional behaviors. Reflexes are relatively rigid and typically do not interface in a rich way with other psychological states-- they do not need to interact with attention or memory, for instance. They just connect sensory inputs to motor outputs (the reality is more complicated, but let's simplify for the sake of the examples). So emotions are more complex than reflexes, they "decouple" stimuli from responses, thus affording much more flexibility [2]. Planned, volitional behavior, on the other hand, is more flexible and more complex than emotions. Emotional behaviors are not like that either -- they don't have that many degrees of freedom. Emotions regulate behavior at a level of complexity intermediate to that of reflexes and volitional behavior [3]. Charles Darwin had a similar notion in mind when he wrote about emotional behaviors, "*Nor can these movements in the dog be explained by acts of volition or necessary instincts, any more than the beaming eyes and smiling cheeks of a man when he meets an old friend*." [1]

In my view, then, emotion states evolved in order to allow us to cope with environmental challenges in a way that is more flexible, predictive, and context-sensitive than are reflexes, but that doesn't yet require the full flexibility of volitional, planned behavior. They evolved to deal with particular, recurring themes in our environment; and because most of the specific sensory features of those themes are highly variable, they also critically involve learning. Broadly, emotions are one solution to determining what is relevant in the world by learning recurring patterns-- themes, if you will. In fact, I think that the patterns that emotions are tuned to are at the level of "core relational themes" [4], even if the specific relational themes that psychological theories have proposed so far may not be the right ones. This is a functional definition of emotion states (I use the term "functional" in the philosophical sense of functionalism, not in the developmental psychological sense. A functionally defined term is defined by what it does rather than by how it is constituted. Consequently animals with very different

brains -- and, in principle, even robots -- could nonetheless be in similar emotion states).

Building on the comments above, we can begin to list the properties that emotion states exhibit, which psychological theories of emotion have often attempted to do, and which my colleague David Anderson and I have also attempted [5]. **Figure 2** summarizes some of these in a provisional list. One universally recognized feature of emotions is that they can be related to one another in a similarity space. The simplest such space has two dimensions (perhaps corresponding to something like valence and arousal), although we will need additional dimensions to capture all the varieties of emotions. Fairly discrete clusters of emotion instances in this space would then correspond to specific emotion categories. Another prominent feature of emotion states is their flexibility, which derives from their "decoupled" nature: they are central states that persist for some time, and so can accumulate a host of contextual information before triggering behavioral decisions. The persistence of an emotion state also permits it to interface in a rich way with the rest of cognition, a major topic of study [6, 7]. Given the multiple causal effects of an emotion state, these need to be coordinated in some way, another feature that is frequently emphasized in psychological theories of emotion (and one that, for me, is strong evidence for a central emotion state that does the coordinating). Finally, emotion states have prepotent control over behavior, a feature similar to historical schemes of an "interrupt" mechanism that can terminate ongoing goal-directed behavior when a sudden environmental challenge is encountered [8]. All of this is of course further complicated in adult humans, since there is some degree of volitional control over emotion states and their expressions. Emotion regulation and strategic/deceptive signaling through emotional expression are perhaps the most distinctively human aspects of emotion.

The features of emotion states sketched in **Figure 2** are relatively domain-general, but in their combination provide clues to the domain-specific roles that guided the evolution of particular emotions. The way that many of the features are engaged relative to specific stimuli and behaviors will demarcate emotion categories. For instance, conditioned taste aversion or Pavlovian fear conditioning both involve learning (by itself a domain-general feature), but they apply to specific kinds of stimuli-- not just any stimulus can be conditioned in this way, and not just any behavior can be produced (only those stimuli, and behaviors, relevant to dealing with threats and to avoiding poisonous foods, in this example). In many cases, the possible functional roles that an adult human emotion can play are enormous, so I believe that we should begin the investigation by identifying the core functional roles that specify the emotion category. This is one reason that developmental and comparative data are essential. They can give us some hints as to what the functional relations are that guided the evolution of the emotion -- this is the ultimate functional story we would want to know: what functions, in our ancestral environment, did an emotion play that resulted in the selection of neural mechanisms to implement that function? Evolutionary psychology tries to provide precisely such functional accounts of emotion states [9], including not only accounts for emotions like fear and disgust, but also for social emotions: these serve functional roles in regulating our social behavior [10]. One exciting approach in

affective neuroscience could be to design experiments that engage functional roles for specific emotions as hypothesized by theories from evolutionary psychology.

I think many of the attributes of emotion states, and the functional conception of them that I am advocating, would put me in the "Basic Emotion" camp, even though I would disagree with many other details of certain Basic Emotion theories. By "basic" I mean "biologically basic" [11], that is, a category defined in an evolutionary sense. I take it that this is also the sense in which psychologists like Ekman [12] have used the term, and the sense in which neurobiologists like Panksepp [13] have used the term (although they posit different sets of basic emotions). Basic Emotion theorists typically have not only a list of the specific features an emotion must satisfy in order to count as basic (e.g., culturally universal, specific physiological profile, etc.) but also supply their inventory of the particular basic emotions (e.g., happiness, fear, anger, etc.). While I think that we can indeed begin to list features (cf. **Figure 2**), I am reluctant to actually name a list of emotions, because I do not think we have enough data yet to do this (especially cross-species data), and because the words for emotions that already exist are likely to be misleading. I also see no reason why a specific emotion category could not also be represented in a dimensional space (discrete and dimensional ways of describing emotions seem complementary).

There is a lot of work to be done in order to figure out the functional role of different emotions, and in order to come up with the best categories and/or dimensions by which to characterize the different emotion states. But I think we also have a lot of data already that points the way for how best to parse emotions. Some of those data are from psychological studies in humans, some from behavioral studies in animals, and some from neuroscience studies in both species. Perhaps the best place to start, if one wanted to begin neuroscientific studies of a specific emotion, would be to pick an emotion on which there is a fair amount of evidence across all these different sources. Emotion categories like anger, fear or disgust seem particularly well suited, for example.

2. How to get confused about emotions.

Affective neuroscience can be confusing when it fails to make distinctions between different aspects of affective processing. The titles of papers and discussions that authors give are often no help here either, since they frequently conflate different meanings of the term "emotion". The most common ambiguity is between "emotion" as conceived above (the functional state) and its conscious experience, conceptualization, or attribution. Generally, when I am in a state of anger, I also feel angry, and I also think about being angry. Those are all interesting processes to study, but they are distinct. I am interested here, in the first instance, in how we should study "emotion states," not "how people can use concepts to think about emotions", or "how people make attributions of emotions," or "how people can speak about emotions". Those are all further interesting questions, and certainly questions that affective neuroscience should investigate, but I don't believe they are the place to start because they don't help us to ground what emotions are supposed to be about in the first place.

Take again the example of my hissing cat. The cat cannot speak about emotions, plausibly also cannot think about nor has a concept of what an emotion is, and it remains unclear how to determine if it would even have a conscious experience of an emotion, whatever one means by that exactly. But it is clear to me that it has emotions (in the above functionally defined sense) -- this is what I attribute to the animal in order to explain and predict its behavior, and indeed it works fairly well most of the time. The emotion states are the internal functional states that produce the behaviors we see. Colloquially, that is what we would say about the cat: it is in a "state of fear". Affective neuroscience investigations of these states would then study what it is in the brain of the hissing cat that causes those behaviors, in response to particular context-dependent stimuli. Of course, there are many more subtle and mixed emotions than these blatant examples, as well as more sustained states that we would typically call moods, and there can be further debate about where to draw the line and say that we would no longer call a state an emotion state. To anchor the investigation, however, I believe we need clear, strong emotion states evoked by ecologically valid stimuli, where "ecologically valid" simply means "experimentally re-creating the functional challenge that we hypothesize engages the emotion under investigation".

Emotion states, then, are not the same as emotion concepts or emotion experiences. By analogy, if I wanted to study people's concepts of planets, I could do psychological or neuroimaging studies to investigate this (I would study what people know and think about planets). But if I want to study planets, I would do astronomy and use a telescope. Just like concepts of planets are not planets, concepts of emotions are not emotions. Emotion states are also not the same as conscious experiences of emotions. In this respect, the usage of the word "emotion" that I am advocating is similar to how we use words like "vision" or "memory" in neuroscience. For the layperson, vision and memory are all about conscious experiences (seeing and remembering). For the scientist they are functionally defined terms, and indeed we now know that both visual processing and memory can be non-conscious (as in blindsight and nondeclarative memory). Our commonsense concepts for most mental-state terms seem to depend on our concept of conscious experience, but I think our scientific concepts for mental-state terms should not. If they did, it becomes problematic how to study the minds of people and animals who cannot use language to tell us about their conscious experiences. Joseph LeDoux has correctly pointed out this problem: if we use the commonsense concept of "fear" when describing animal neuroscience, we risk confusing this with the attribution of a conscious experience of fear. LeDoux concludes from this that we should stop using words like "fear" or "emotion" when doing animal neuroscience [14], but I think there is a simpler solution: do the same thing we as scientists do when we study vision or memory. Use a scientific concept of "emotion states" that is not based on conscious experience.

To summarize how people get confused about what is meant by "emotion": there are distinctions between the functional emotion state ("the emotion state"), its conscious experience ("the experience of the emotion"), our ability to attribute emotions to others and to animals ("attribution of emotion"; "emotion perception"), our ability to think and

talk about emotion (“conceptualizing emotion”), and the behaviors caused by an emotion state (“the expression of emotions”, “emotional reactions”) (**Table 1**). I think emotions are first and foremost about the first of these, and all the others are derivative (but no less interesting to study).

3. Dissociating emotion states from emotion concepts: An example from neuroscience.

An example of how emotion states can be separated from conceptual knowledge of emotions comes from classical cognitive neuropsychology, the use of neuroscience data in order to help make distinctions by showing dissociations [15]. The famous patient S.M. [16], who has bilateral lesions to her amygdala, shows a dissociation with respect to fear that is about as good as dissociations get. She can laugh, she can cry, and she can endorse feelings of happiness and sadness and most other emotions. But she does not show any of the effects of a state of fear that we would normally use in order to attribute fear [17]. She does not show normal avoidant behaviors to threatening situations, she does not show autonomic responses or give subjective ratings of fear to normally fear-inducing stimuli, and she fails to show learning based on unconditioned fear in Pavlovian fear conditioning. A subset of the same deficits (minus the subjective ratings and with simpler behaviors) is seen in animals with bilateral amygdala lesions [18].

Despite the complete failure to induce a state of fear from any external stimuli, S.M. can tell us a lot about fear. She has read books and watched movies in which fear occurs, she has spoken with other people about fear, and she even has autobiographical memories of feeling afraid as a child (plausibly before her amygdala was fully lesioned) [16, 17]. Consequently, she has accumulated an impressive store of semantic knowledge about fear, so much so that we should say that she has a concept of fear. She can tell you that people who are afraid scream and run away; she can tell you that being chased by a bear would make you afraid; she can use the word “fear” appropriately in conversation. But she herself does not instantiate the state of fear, even though she has so much conceptual knowledge about fear. Just having the concept of fear is typically insufficient for the state of fear. In fact, it is extraordinarily difficult to induce an emotion state by just activating the concept of an emotion. If it were easy, depressed people would just need to think about being happy and they would be happy.

In humans, the routes by which a state of fear can be induced are of course considerably more varied than in other animals, and include memories and imaginings in addition to actual occurent sensory stimuli. Indeed, if I think hard about situations in which I would be afraid, I feel a little bit afraid. So conceptual representations of emotions do have some effect at least on the conscious experience of emotions, and presumably on the emotion state as well. Conversely, being in an emotion state typically also causes conceptual representations of emotions. If you are in a state of fear, you typically also think about fear and believe you are in a state of fear. So another important challenge for affective neuroscience is to detail the causal

interactions between emotion states, emotion experiences, and emotion concepts: usually, all three occur together.

There is a final important dissociation that patient S.M. showed us. As with the case of memory, where H.M. showed us how declarative and nondeclarative memory can be dissociated, S.M. has also shown us that fear to external stimuli, and panic to interoceptive stimuli, can be dissociated. S.M. does panic if she feels like she is suffocating (elicited in the experiment through inhaling carbon dioxide [19]). This was a very interesting advance, and showed us that our scientific concept of fear needs finer distinctions, just like H.M. and many studies since then have given us a more fine-grained taxonomy of memory. So the case of S.M. shows us three different dissociations relevant to affective neuroscience: that the state of fear can be dissociated from other emotion states; that the state of fear can be dissociated from its concept; and that there are varieties of fear, to which we may want to give separate names (anxiety, fear, and panic have all been used already, and there may be additional varieties that function with respect to specific types of threat [20]).

4. A framework for neuroimaging studies of emotion states: systems, hierarchy and topography.

Explaining how an emotion state is implemented in the brain requires us to explain which structures, at which point in time, implement particular computations. That makes it nonsensical to ask if "fear is in the amygdala", for example, since the state is distributed in both space and time. Nonetheless, we can say that the amygdala is one component of the neural system for fear, and moreover a necessary component. At the coarsest level, we know there's something happening in somebody's brain for the several seconds or minutes they are in a state of fear. At the most microscopic level, we know there are causal events, each at the millisecond scale, across billions of neurons. The first description is too low dimensional; the second is too high dimensional. So the challenge to the neuroscientist is: can you find something useful in the middle, something at the level of neural systems, that eventually allows us to understand how emotions link stimuli to behavior (and other cognitive states).

Figure 2 can motivate initial hypotheses here. As already noted, one prominent and universally acknowledged feature of emotions is that they have a similarity structure. Anger and fear are more similar than anger and happiness, for example. Similarity relationships are often partly captured in a two-dimensional space of valence and arousal. These facts motivate the hypothesis that there should be topography in how emotions are instantiated in the brain. Indeed, studies in rodents have argued for a topography in the nucleus accumbens that maps the dimension of valence [21], at least with respect to feelings. One challenge in discovering topography is that we will need a better description of the dimensional space that defines similarity relations among emotion states. On the other hand, semantic knowledge for emotions has indeed been mapped, at least across cortex, and can be compared to semantic knowledge of many other concepts for which we have words [22]. Another study [23] was notable for

comparing similarity amongst neural representations with conceptual similarity in how people rate emotions according to several popular psychological theories (arousal/valence; basic emotions; appraisal dimensions). That study [23] found that attributions about other people's emotions that we make from thinking about the situations in which people find themselves activate representations in a system of brain structures known to be involved in mental state attribution more generally (such as the temporoparietal junction, dorsomedial prefrontal cortex, and precuneus); the best match between neural and psychological similarity structures held for ratings on appraisal dimensions. These recent examples use techniques that would likely be useful also to investigate how emotion states (functionally defined) might be topographically represented in the brain (voxel-based modeling [22], and representational dissimilarity analysis [23], respectively).

Another reasonable hypothesis derives from the hierarchical coordination that emotion states achieve. In this respect, they feature what Tinbergen already observed in the fixed action patterns many animals exhibited [24]. Some neurobiological studies, especially in rodents, have produced very detailed knowledge of certain components of an emotion state; for instance, optogenetic activation of specific neuronal populations in the hypothalamus can produce directed aggressive behavior in mice [25]. The next question is: what are the inputs to these hypothalamic neurons that would normally orchestrate this behavioral component? The coordination amongst many components requires a hierarchical control of sorts, and we could test whether this is accomplished by yet a separate component, or whether it arises from network dynamics amongst all the pieces. This is basically what systems neuroscience is already doing: studying specific components of an emotion, and trying to figure out how they are connected to produce all the features of an emotion. Such systems-level studies of emotions in animals try to choose ecologically valid situations to induce emotion states like fear, or try to dissect specific components of such states, like fear learning in Pavlovian fear conditioning.

So we have some promising examples of topography in human fMRI studies that were not about actual emotion states but instead about concepts; and from the components of actual emotion states but studied in animals. By contrast, there have been very few neuroimaging studies in humans that have attempted to actually induce real emotions in human subjects, and fewer still that have attempted to dissociate them from experiences or conceptual processing of emotions. Nonetheless, there have been a handful of important imaging studies, ranging from early ones with PET [26] to later ones with fMRI [27, 28], that derive conclusions about particular brain structures involved in processing emotions from trying to induce actual specific emotion states (through autobiographical recall of emotional events, fear of electric shock, or with innately emotion-inducing stimuli like tarantulas, respectively). Those studies have focused on brain structures mostly distinct from the cortical regions emphasized in studies of semantic representation or mental state attribution. They have instead emphasized subcortical structures like the amygdala, hypothalamus, and periaqueductal gray, much as have the studies of emotion states in animals (as well as additional cortical regions such as ventromedial prefrontal cortex and insula). All of

these regions have also been noted in human and animal lesion studies, and have been the material for several influential neurobiological theories of emotion [29, 30].

It is worth noting that our knowledge from fMRI studies of the neural regions encoding information about aspects of emotions has only emerged in the last few years. Initial meta-analyses [31] searching for "basic" emotion representations had little success, no doubt in good part because the studies used for the meta-analysis were underpowered, univariate, and mixed many different aspects of "emotion". That has changed with general increase in sensitivity of fMRI studies and the widespread adoption of multivariate analyses, developments that produced evidence for dimensional components of core affect like valence [32], and, more recently, indeed found evidence for basic emotion representations [33], conclusions replicated also in meta-analyses of studies on emotion categories [34]. The rarity of human studies that have actually produced strong emotion states, and the frequent conflation of emotion states, concepts, and feelings, all suggest that it is far too early to say much about emotion states on the basis of neuroimaging data. On the other hand, there is a wealth of data from animal studies, and there are lesion dissociations in humans, both of which help to motivate strong hypotheses about where to look in the brain, once we figure out how to design good neuroimaging studies of emotion states.

So to what extent can we in fact dissociate the functional emotion state from emotional experience, labeling, or concepts? We could probably minimize the latter two under experimental conditions that prevent reflective processing, or by imaging children or patients with certain kinds of brain damage. I don't know how to dissociate emotional experience from the functional emotion state, but that is a problem faced also by all studies that want to isolate the neural correlates of conscious experience [35]. A principal challenge will be to construct ecologically meaningful situations that can induce strong and well-characterized emotion states in the scanner environment. Specific hypotheses about the functional roles that particular emotions play, perhaps informed by work in evolutionary psychology and ecology, would be needed to design the best experiments. The ideal (difficult) project would design a series of experiments across different species to study, for instance, how the induction of fear across rodents, monkeys, and humans might engage both overlapping neural systems as well as components unique to a particular species. More realistically, we could design human neuroimaging studies that contrast different emotions (or attempt to discover distinguishable subtypes of what we currently think of as one category of emotion), while also varying the level of associated conceptual processing. This would still generally produce results that speak equally to emotion states and emotion experiences, but one cannot dissociate everything at the outset. The recommendations for the affective neuroscientist using fMRI are threefold: (1) partly disentangle the neural correlates of emotions from all the other processing with which they interact; (2) carefully distinguish what aspect of "emotion" it is you are investigating (states, experiences, concepts); (3) construct hypotheses derived from knowledge of the functional features of emotions (**Figure 2**) and investigate these with the most sensitive neuroimaging methods.

As an affective neuroscientist, I would want a framework for investigating emotion that lets me investigate emotions not only in healthy adult humans, but also in rats, in people who cannot talk, and in people who are deluded about what emotion they have. I would even want to be able to say something to engineers who might want to construct robots that have emotions. An operationalization of emotions as functional states lets me do all of these, whereas a focus on emotion experiences, or emotion concepts, does not. Again, all useages of the word “emotion” are interesting to study, and it may well turn out that emotion states, emotion experiences, emotion concepts, and emotion attributions are all related in interesting ways, and that they share neural substrates. But that needs to be an empirical result, not something we conflate in our research program at the start.

Acknowledgements: I thank Laura Harrison, Bob Spunt, Damian Stanley, Julien Dubois, David Anderson, and Ajay Satpute for helpful comments on the manuscript.

Table 1. Some examples of different aspects of emotion investigated in affective neuroscience, and my opinion about how central they are to a functionally defined emotion state.

Aspects of processing that are central to an emotion state.

- emotion-cognition interactions [6, 7]
- emotional learning and memory [36]
- eliciting strong emotions with ecological stimuli [19, 28]

Aspects of processing that are less central to an emotion state.

- perceiving emotional social signals (emotion perception) [37]
- inferring emotions in other people (social inference, theory of mind) [23, 38]
- semantic processing about emotions (concepts) [22]
- lexical processing about emotions (words) [39]

Figure 1. Rate your position. Which applies to emotions? Indicated are **my own**, my take on **Lisa Barrett's** [40, 41] and my take on **Jaak Panksepp's** [13], to provide three different views (any errors are of course mine). Many of the terms have unclear meanings, and the figure is intended only to give a rough starting point for discussions, not to quantify theoretical frameworks. Lisa saw a prior version of this figure and sent some corrections to my original take on her view. My original depictions of her positions are indicated by circles; the corrected positions from Lisa are denoted by triangles.

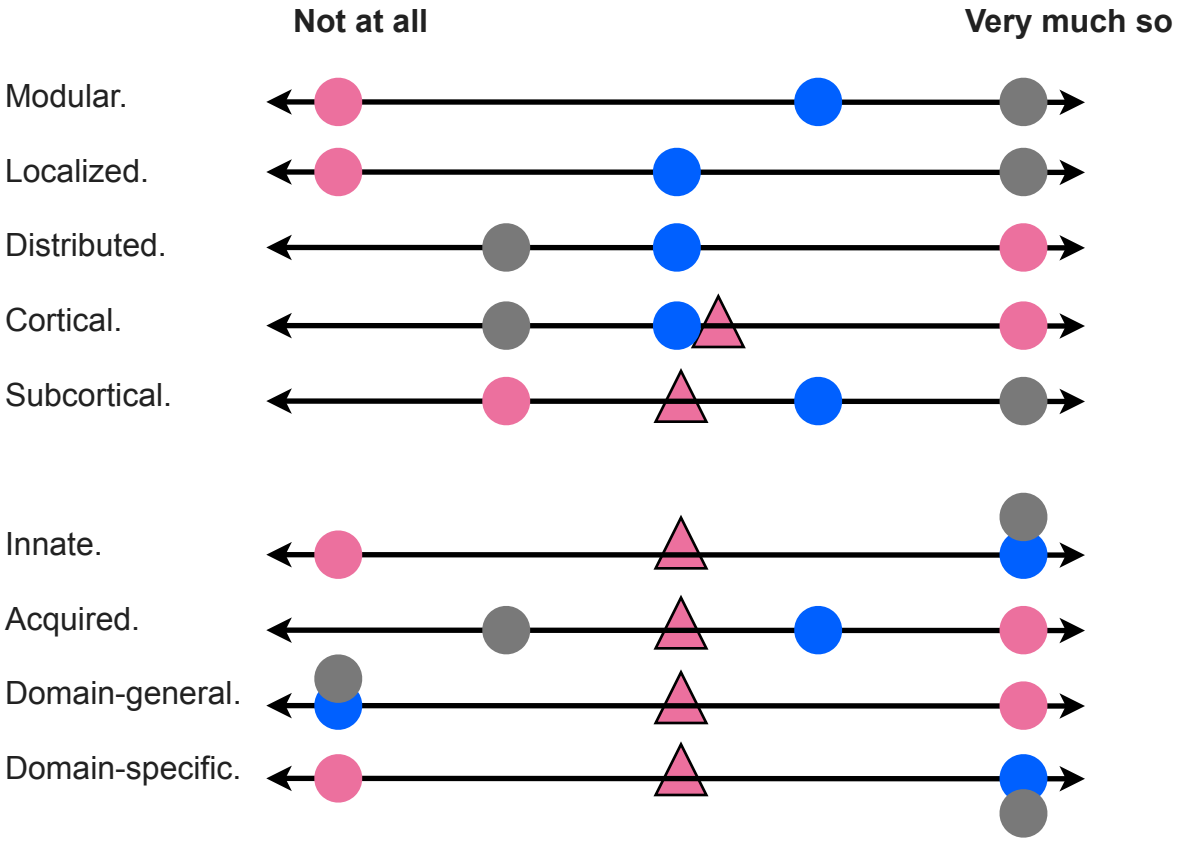
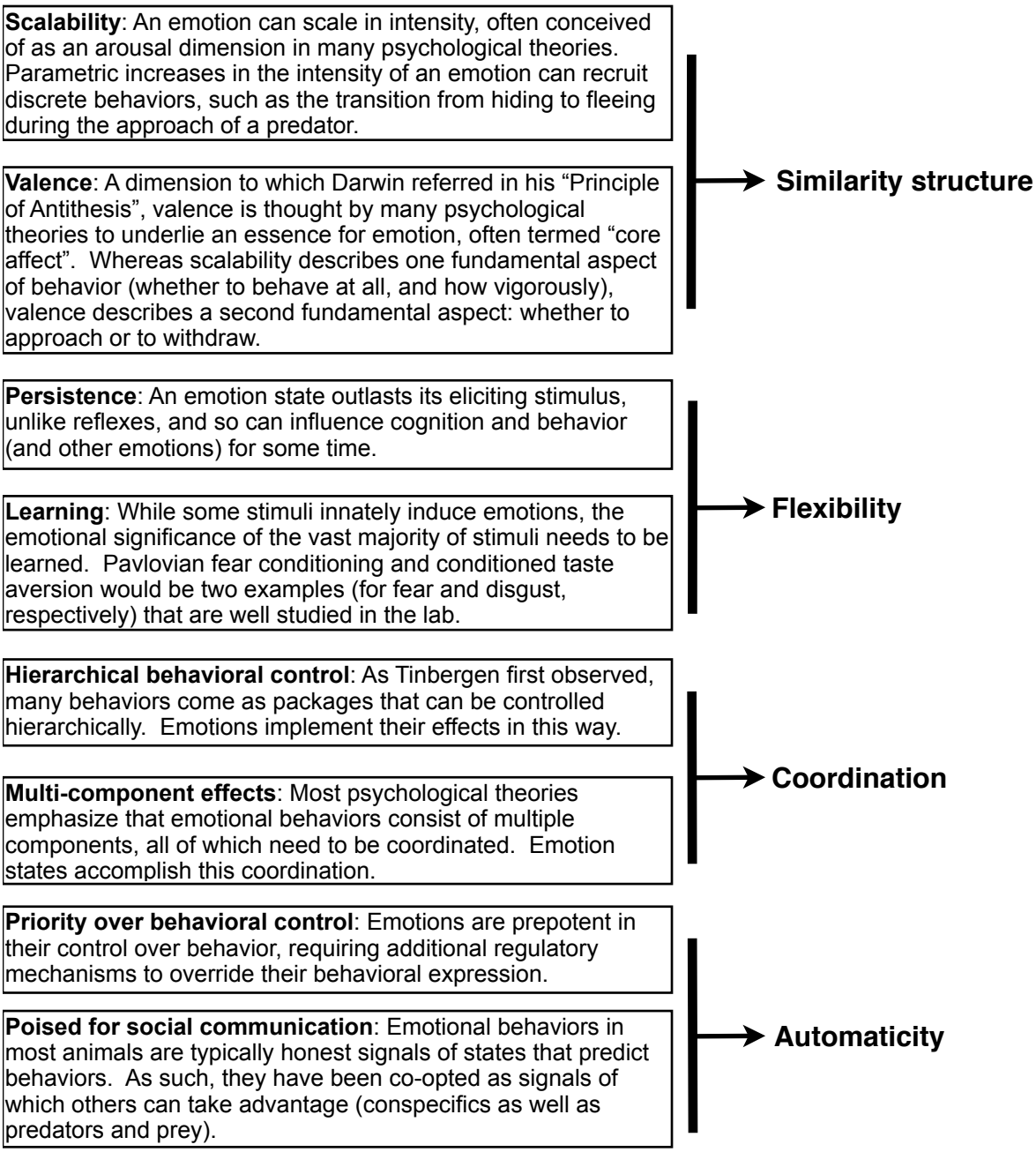


Figure 2. Features of an emotion. These features begin to describe functional properties of emotions, some of which may help define dimensions or categories of emotions in a mature theory of emotion. Several of these are modified from [5].



Bibliography

1. Darwin, C. (1872/1965). The Expression of the Emotions in Man and Animals, (Chicago: University of Chicago Press).
2. Scherer, K.R. (1994). Emotions serve to decouple stimulus and response. In The Nature of Emotion, P. Ekman and R.J. Davidson, eds. (New York: Oxford University Press), pp. 127-130.
3. Adolphs, R. (in press). Emotions are functional states that cause feelings and behavior. In The Nature of Emotion, 2nd Edition, R.J. Davidson, A. Shackman, A. Fox and R. Lapate, eds. (New York: Oxford University Press).
4. Lazarus, R.S. (1991). Emotion and Adaptation, (New York: Oxford University Press).
5. Anderson, D.J., and Adolphs, R. (2014). A framework for investigating emotion across species. Cell *in press*.
6. Okon-Singer, H., Hendler, T., Pessoa, L., and Shackman, A. (2015). The neurobiology of emotion-cognition interactions: fundamental questions and strategies for future research. Frontiers in Human Neuroscience *9*.
7. Pessoa, L. (2013). The Cognitive-Emotional Brain: From Interactions to Integration, (Cambridge, MA: MIT Press).
8. Simon, H.A. (1967). Motivational and emotional controls of cognition. Psychological Review *74*, 29-39.
9. Tooby, J., and Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In Handbook of Emotions, M. Lewis, J.M. Haviland-Jones and L.F. Barrett, eds. (New York: Guilford Press), pp. 114-137.
10. Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S., and Halperin, E. (2016). Shame closely tracks the threat of devaluation by others, even across cultures. PNAS *www.pnas.org/cgi/doi/10.1073/pnas.1514699113*.
11. Ortony, A., and Turner, T.J. (1990). What's basic about basic emotions? Psychological Review *97*, 315-331.
12. Ekman, P. (1999). Basic emotions. In Handbook of Cognition and Emotion, T. Dalgleish and M. Power, eds. (Chichester, UK: John Wiley), pp. 45-60.
13. Panksepp, J. (1998). Affective Neuroscience, (New York: Oxford University Press).
14. LeDoux, J. (2012). Rethinking the emotional brain. Neuron *73*, 653-676.
15. Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. Brain and Cognition *5*, 41-66.
16. Feinstein, J.S., Adolphs, R., and Tranel, D. (2016). A tale of survival from the world of patient S.M. In Living without an Amygdala, D.G. Amaral and R. Adolphs, eds. (New York: Guilford Press).
17. Feinstein, J.S., Adolphs, R., Damasio, A., and Tranel, D. (2011). The human amygdala and the induction and experience of fear. Current Biology *21*, 34-38.
18. Amaral, D.G., and Adolphs, R. eds. (2016). Living without an amygdala (New York: Guilford Press).

19. Feinstein, J.S., Buzza, C., Hurlemann, R., Follmer, R.L., Dahdaleh, N.S., Coryell, W.H., Welsh, M.J., Tranel, D., and Wemmie, J.A. (2013). Fear and panic in humans with bilateral amygdala damage. Nat Neurosci *16*, 270-272.
20. Silva, B.A., Mattucci, C., Krzywkowski, P., Murana, E., Illarionova, A., Grinevich, V., Canteras, N.S., Ragozzino, D., and Gross, C.T. (2013). Independent hypothalamic circuits for social and predator fear. Nature Neuroscience *doi: 10.1038/nn.3573*.
21. Berridge, K.C., and Kringelbach, M.L. (2013). Neuroscience of affect: brain mechanisms of pleasure and displeasure. Current Opinion in Neurobiology *23*, 294-303.
22. Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., and Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature *532*, 453-458.
23. Skerry, A.E., and Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. Current Biology *25*, 1-10.
24. Tinbergen, N. (1950). The hierarchical organization of nervous mechanisms underlying instinctive behaviour. In Physiological Mechanisms in Animal Behaviour, Volume IV. (New York, NY: Academic Press, Inc.), pp. 305-312.
25. Lin, D., Boyle, M.P., Dollar, P., Lee, H., Lein, E.S., Perona, P., and Anderson, D.J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. Nature *470*, 221-226.
26. Damasio, A.R., Grabowski, T.J., Bechara, A., Damasio, H., Ponto, L.L.B., Parvizi, J., and Hichwa, R.D. (2000). Feeling emotions: subcortical and cortical brain activity during the experience of self-generated emotions. Nature Neuroscience *3*, 1049-1056.
27. Mobbs, D., Petrovic, P., Marchant, J.L., Hassabis, D., Weiskopf, N., Seymour, B., Dolan, R., and Frith, C. (2007). When fear is near: threat imminence elicits prefrontal-periadqueductal gray shifts in humans. Science *317*, 1079-1083.
28. Mobbs, D., Yu, R., Rowe, J.B., Eich, H., FeldmanHall, O., and Dalgleish, T. (2010). Neural activity associated with monitoring the oscillating threat value of a tarantula. PNAS *early edition*, www.pnas.org/cgi/doi/10.1073/pnas.1009076107.
29. Damasio, A. (2003). Looking for Spinoza: Joy, Sorrow, and the Feeling Brain, (Orlando, Florida: Harcourt, Inc.).
30. Craig, A.D. (2008). Interoception and Emotion: A Neuroanatomical Perspective. In Handbook of Emotions, 3rd Edition, J.M.H.-J. M. Lewis, L. Feldman-Barrett, ed. (New York: Guilford Press), pp. 272-288.
31. Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., and Feldman Barrett, L. (2012). The brain basis of emotion: a meta-analytic review. Behavioral and Brain Sciences *35*, 121-143.
32. Chikazoe, J., Lee, D.H., Kriegeskorte, N., and Anderson, A.K. (2014). Population coding of affect across stimuli, modalities, and individuals. Nature Neuroscience *17*, 1114-1122.
33. Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., and Nummenmaa, L. (2015). Discrete neural signatures of basic emotions. Cerebral Cortex *doi: 10.1093/cercor/bhv086*.

34. Wager, T.D., Kang, J., Johnson, T.D., Nichols, T.E., Satpute, A.B., and Barrett, L.F. (2015). A Bayesian model of category-specific emotional brain responses. PloS Computational Biology *DOI:10.1371/journal.pcbi.1004066*.
35. Block, N. (2015). Consciousness, big science, and conceptual clarity. In The Future of the Brain, G. Marcus and J. Freeman, eds. (Princeton, NJ: Princeton University Press).
36. Phelps, E.A. (2006). Emotion and cognition: insights from studies of the human amygdala. Annual Review of Psychology *57*, 27-53.
37. Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature *372*, 669-672.
38. Spunt, R.P., and Lieberman, M.D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. Neuroimage *59*, 3050-3059.
39. Rapcsak, S.Z., Kaszniak, A.W., and Rubens, A.B. (1989). Anomia for facial expressions: evidence for a category specific visual-verbal disconnection syndrome. Neuropsychologia *27*, 1031-1041.
40. Barrett, L.F. (2006). Solving the emotion paradox: categorization and the experience of emotion. Personality and Social Psychology Review *10*, 20-46.
41. Wilson-Mendenhall, C.D., Barrett, L.F., Simmons, W.K., and Barsalou, L.W. (2011). Grounding emotion in situated conceptualization. Neuropsychologia *49*, 1105-1127.